DOCUMENT RESUME

ED 116 938 SE 020 141

AUTHOR	Begle, E. G.
TITLE	Ability Grouping for Mathematics Instruction. A Review of the Empirical Literature. SMESG Working Paper No. 17.
INSTITUTION	Stanford Univ., Calif. Stanford Mathematics Education Study Group.
REPORT NO	SMESG-WP-17
PUB DATE	Dec 75
NOTE	55p.; Occasional marginal legibility; Best copy available
EDRS PRICE	MF-$0.76 HC-$3.32 Plus Postage
DESCRIPTORS	*Ability Grouping; Achievement; *Annotated Bibliographies; Attitudes; Elementary School Mathematics; Elementary Secondary Education; Instruction; *Literature Reviews; *Mathematics Education; Research; *Research Reviews (Publications); Secondary School Mathematics

ABSTRACT

This paper examines the research literature concerning the effects of homogeneous vs. heterogeneous ability grouping on the mathematical achievement and attitudes of students. The studies examined concern all grade levels from 1 through 12, and relate to grouping on the basis of general ability and on the basis of prior mathematical achievement. Differential effectiveness of grouping patterns on students of high, middle, and low ability is examined. Grouping within individual classrooms is also discussed. On the basis of the analysis of the literature, three recommendations are made: (1) school systems should be encouraged to experiment with homogeneous grouping; (2) teachers should be encouraged to experiment with grouping students within the classroom; and (3) research on team-teaching should be done. The second part of this paper provides an annotated bibliography. Seventy-seven research reports related to ability grouping are annotated. For each report bibliographic information, a summary of the research methods, and a list of findings are provided. (SD)

STANFORD
MATHEMATICS
EDUCATION
STUDY GROUP

Working Paper No. 17 December, 1975

Ability Grouping for Mathematics Instruction. A Review of the Empirical Literature.

By: E. G. Begle

Preface

This is one of a series of reveiws of empirical studies of various aspects of mathematics education.

This review has two parts. Part B is an annotated bibliography of reports published since the beginning of 1960. The annotation for each report specifies the grade levels at which the study carried out, the procedures used in the study, and the kinds of information collected. The results of the study are then summarized.

In many of these studies other subject matters, in addition to mathematics, were investigated. I do not record the findings with respect to these except occasionally to note that the pattern of findings for the other subjects was different from that for mathematics.

The quality of the design, execution, and analysis of these studies varies greatly. In general this quality has increased with time. However, comments along these lines are not included in the annotations. Due to the great variations across this country in students, teachers and schools, no study or small set of studies, even if of the highest quality, could definitively determine the value of ability grouping.

However, it seems likely that if ability grouping had indeed been a strong educational variable then that fact would have clearly emerged provided enough studies were carried out, even if each of them was less than perfect in design, execution, or analysis. Consequently, a strong effort was made to locate all reports on ability grouping in which measures of mathematics achievement were studied.

Any reader knowing of references I have missed is urged to bring them to my attention.

I have not attempted to provide a complete list of other reviews and discussions, but a short list of such publications is provided at the end of the bibliography.

These annotations form the bulk of this report. However, Part A is an attempt to use these annotations to provide tentative answers to some of the questions most asked about ability grouping and to provide some suggestions to teachers, school officials, and educational researchers about the use of ability grouping.

PART A

Introduction

"Ability grouping" is a term used to designate any procedure for assigning students to classes in such a way that the range of abilities within any class is less than it would be if the students had been randomly assigned. Since ability grouping is an attempt to make each class as homogeneous as possible with respect to some ability the phrase "homogeneous grouping" is often used as a synonym for "ability grouping". Similarily the phrase "heterogeneous grouping" is synonymous with "random grouping".

A wide variety of abilities have been used in forming homogeneous classes. These abilities fall into two classes. In the first are such general abilities as are measured by IQ, overall grade point average, etc. In distinction to these are specific abilities such as grade point average in a specific subject matter (mathematics in this report) or scores on a standardized subject matter achievement test.

Grouping by a general ability is usually used in the elementary schools and grouping by a specific ability at the secondary level.

Proponents of ability grouping claim that a teacher can take account of individual differences among students more easily when the range of these differences is small than when it is large and that, consequently, teachers will be more effective and students will achieve better in homogeneous than in heterogeneous classes.

Opponents of ability grouping, however, claim that it can have unfortunate affective results and that it causes low ability students to be stigmatized and to develop a poor self-concept.

The evidence bearing on these conflicting points of view is summarized in Part B of this report. Here we list some specific questions and then outline the relevant information.

Does Homogeneous Grouping Result in Improved Mathematics Achievement?

I separate the evidence bearing on this question into two sets. The first set consists of those studies in which the grouping was carried out on the basis of some measure of general ability, such as IQ or overall grade point average.

The second set of studies are those in which the grouping was based on mathematics achievement, although quite often other criteria, such as teacher recommendations, were also taken into account.

The studies in each of these two sets are divided into four categories. In the first category are placed those studies in which the results include at least one significant difference in favor of homogeneous grouping and no significant difference in the other direction. The second category consists

of those in which there is at least one significant difference in favor of heterogeneous grouping and no significant differences in favor of homogeneous grouping.

The third category consists of those studies with mixed results, i.e. at least one significant difference in favor of homogeneous grouping and at least one other significant difference in favor of heterogeneous grouping. The final category consists of those studies in which no significant difference at all were discovered.

Those studies, identified by author and year, in the first set are listed by category in Table I and the studies in the second set in Table II.

It is important to notice that the number of studies listed in the last category (NSD) is a poor indicator of the total number of non-significant differences reported. Many of the studies in the first three categories included findings of non-significant differences as well as significant ones. It is probably the case that the total number of non-significant differences is substantially greater than the number of significant ones. A look in Part B at the summary of Borg's study will make this clear.

An inspection of Table I does not provide clear cut evidence either for or against homogeneous grouping. A total of 12 studies (categories 1 and 3) reported significant differences in favor of homogeneous grouping but a total of 9 (categories 2 and 3) reported significant differences in favor of heterogeneous grouping. Eight studies reported no significant differences at all, but as pointed out in the preceeding paragraph many other non-significant differences appear in the first three categories.

Consequently, one tentative answer to our first question is that grouping students homogeneously by IQ or general ability is not a very reliable way of improving mathematics achievement.

Turning now to Table II we see a somewhat different picture. Sixteen studies produced significant differences in favor of homogeneous grouping and only 8 produced significant differences in the opposite direction. Once again, however, there were a substantial number of non-significant differences.

Some light on the differences between Tables I and II comes from a study by Balow and Curtin (1966). Half the students in the school were sorted by IQ into three homogeneous sets. The other half formed a heterogeneous group. The variance on achievement for the four sets were not significantly different.

Along the same lines, Balow (1964) administered achievement tests to seventh graders who had been homogeneously grouped on the basis of sixth grade achievement. There was a great deal of overlap on seventh grade achievement scores between the three "homogeneous" classes.

TABLE I

Effects of Grouping on the Basis of General Ability

Grade Levels	Homogeneous Grouping Favored	Heterogeneous Grouping Favored	Mixed Results	No Significant Difference
2-5	Karnes et al (63)			
3-5				Hofset (70)
3-5				Hofset (71)
3-6			Barker Lunn (70)	
3-6	West & Sievers (60)			
3-9	Alam (68)			
4-5	Cluff (64)			
4-6		Bouri & Barker Lunn (69)		
4-6		Daniels (61)		
4-6				Morgenstern (63)
4-7			Borg IV* (64)	
4-9	Dahllof (71)			
4-9				Svensson (62)
5,6		Goldberg et al (66)		
6	Balow & Ruddell (63)			
6				Becker (63)
7,8		Peterson (66)		
8			Ferri (71)	
8			Wilcox (63)	
8-12	Enzmann (61)			
9	Campbell & Knill (65)			
9				Chiotti (61)
10				French (60)
10				Lovell (60)
10		Sawchuck & Black (61)		

*Borg studied five separate sets of students. Borg IV refers to the students who started the study in grade four.

Note: The underlined entries in this and the next table refer to studies in which the curriculum was adjusted to the ability levels of the students. This matter is discussed below.

TABLE II.

Grouping on the Basis of Mathematical Achievement

Grade Level	Homogeneous Grouping Favored	Heterogeneous Grouping Favored	Mixed Results	No Significant Differences
1				Johnston (73)
2-6			Gilbert & McKie (69)	
3			Wardrop et al (67)	
4		Koontz (61)		
4-6		Davis & Tracy (63)		
4-6	Echternoch & Gordon (62)			
4-6	McLaughlin (61)			
4-6	Provus (60)			
6	Balow & Ruddell (63)			
6	Pinney (61)			
6-9			Borg VI (64)	
7				Bachman (68)
7,8	Adamson (71)			
7,8				Mahler (61)
7,8				Mikkelson (62)
7,8				Millman & Johnson (64)
7-10			Borg VII (64)	
8	Sommers (60)			
8-11			Borg VII (64)	
9				Bailey (67)
9	Campbell & Knill (65)			
9				Chiotti (61)
9	Frye (62)			
9-12	Borg IX (64)			
13			Schrank (70)	

A comparison of Tables I and II makes it clear that if students are to be grouped homogeneously, then the grouping criterion should be previous mathematics achievement rather than general ability. However, even this method of grouping is not a dependable way of improving mathematics achievement. With this method, it is more likely that achievement will improve than it will decrease but it is still more likely that there will be no significant difference.

Does Homogeneous Grouping Have Different Effects at Different Ability Levels?

Not all of the studies listed in Part B reported mathematics achievement results for low, medium, and high ability students separately. Tables III A and B, IV A and B, and V A and B identify studies that did. Table III A identifies those studies in which homogeneous grouping was based on general ability and in which separate reports for low ability students were made. Table III B identifies those studies in which grouping was based on mathematics achievement and in which separate reports for low ability students were provided.

Tables IV A and B provide corresponding information for middle ability students and Tables V A and B provide information for high ability students.

It should be noted however that definitions of low, middle and high ability are not uniform across these studies.

The patterns displayed in Tables III A and III B suggest rather clearly that homogeneous grouping is not a very promising practice for low ability students. The pattern in Table IV A suggests that homogeneous grouping should not be used with middle ability students when the grouping criterion is general ability. Table IV B suggests that even when the grouping criterion is mathematics achievement, homogeneous grouping is not a promising practice for these students.

However, Tables V A and B show an entirely different picture and argue strongly that homogeneous grouping should be used for high ability students.

Is it Helpful to Students of one Ability Level to Have Students of Another Ability Level in the Same Class?

It is sometimes stated that high ability students are held back by the presence in the same classroom of low ability students. Similarily, it is argued by some that low ability students need the stimulation provided by high ability students. Unfortunately, few of the studies listed in Part B attempted to directly answer this question.

Bailey (1967) reported that, for high ability students, it made no difference whether there were students in the class at lower ability levels or not. Both Mahler (1961) and Mikkelson (1962) report the same findings.

Goldberg et al (1966) reported that it made no difference to students in the middle ability levels whether or not there were high ability level students in the class. However, they also reported that for lowest ability students it was better to have high ability students in class than not to have them.

TABLE III A

Effects of Ability Grouping for Low Ability Students

Grade Level(s)	Homogeneous Grouping Favored	Heterogeneous Grouping Favored	Mixed Results	No Significant Differences
4-7				Borg IV (64)
5-9				Svensson (62)
7,8		Peterson (66)		
9	Campbell & Knill (65)			

TABLE III B

Effects of Ability Grouping for Low Ability Students

Grade Level(s)	Homogeneous Grouping Favored	Heterogeneous Grouping Favored	Mixed Results	No Significant Differences
3		Wardrop et al (67)		
4		Koontz (61)		
4-6				Provus (60)
6	Dewar (61)			
6-9		Borg VI (64)		
7-10		Borg VII (64)		
8-11	Borg VIII (64)			
9-12	Borg IX (64)			

TABLE IV A

Effects of Ability Grouping for Middle Ability Students

Grade Level(s)	Homogeneous Grouping Favored	Heterogeneous Grouping Favored	Mixed Results	No Significant Differences
4-7		Borg IV (64)		
5,6		Goldberg et al (66)		
7,8		Peterson (66)		
9				Campbell & Knill (65)

TABLE IV B

Effects of Ability Grouping for Middle Ability Students

Grade Level(s)	Homogeneous Grouping Favored	Heterogeneous Grouping Favored	Mixed Results	No Significant Differences
3	Wardrop et al (67)			
4		Koontz (61)		
4-6				Provus (60)
6				Dewar (61)
6-9	Borg VI (64)			
7-10		Borg VII (64)		
8-11)		Borg VIII (64)		
9				Bailey (67)
9-12	Borg IX (64)			

TABLE V A

Effects of Ability Grouping for High Ability Students

Grade Level(s)	Homogeneous Grouping Favored	Heterogeneous Grouping Favored	Mixed Results	No Significant Differences
2-5	Karnes et al (63)			
4-7	Borg IV (64)			
5-8	Dahllof (71)			
5-8				Svensson (62)
6	Balow & Ruddell (63)			
8	Wilcox (63)			
9	Campbell & Knill (65)			

TABLE V B

Grade Level(s)	Homogeneous Grouping Favored	Heterogeneous Grouping Favored	Mixed Results	No Sifnificant Differences
3				Wardrop et al (67
4		Koontz (61)		
4-6	Provus (60)			
6	Dewar (61)			
6-9			Borg VI (64)	
7-10	Borg VII (64)			
8-11	Borg VIII (64)			
9				Bailey (67)
9-12	Borg IX (64)			

They also reported that for high and middle ability students it was better not to have, lowest ability level students in the class.

Hofset (1970) reported that it made no difference whether there were few or many high ability students in the class and in Hofset (1971) it was reported that it made no difference whether there were few or many low ability students in the class.

Mikkelson (1962) reported that students not in the high ability level gained about the same whether or not there were high ability students in the class.

In summary, there is little evidence to indicate that students at one ability level benefit from the presence of students at other ability levels.

Can the Large Number of Non-Significant Differences in Studies of Ability Grouping be Explained by Failure to Adjust the Curriculum to the Ability Levels of the Students?

There is not enough evidence to be able to answer this question. Only fourteen of the studies listed in Tables I and II report specifically that the curriculum used was modified to take into account the ability levels of the students. The distribution of these between the four categories is similar to the overall distribution. However, in all of the studies mentioned in the next section, the curriculum was differentiated.

Wilcox (1963) observed that there was very little curriculum variation to take advantage of such grouping as he found in his study.

Mikkelson (1962) states that the curriculum was not differentiated in his study, but he compared homogeneous and heterogeneous grouping only for high ability students.

Are there any other grouping methods?

There have been a number of studies of within class grouping. In these studies each of the experimental classes were divided into subsets, usually three, homogeneous in ability. The teacher would work with one subset at a time while students in the others would work individually or in small groups. It is to be noted that when such a procedure is used each student receives only one third as much direct attention from the teacher as would be the case in the usual classroom, whether homogeneously grouped or not.

There are two variations in the case of within class grouping. In one variation the grouping is established at the beginning of the school year and not changed thereafter. Table VI specifies those studies using this procedure and sorts them into the four categories used in previous Tables. In view of the teacher time allotments mentioned above the results are remarkably good.

It is curious to note that all but one of these studies were done with elementary school students.

TABLE VI

Effects of Within Class Grouping

Grade Level(s)	Homogeneous Grouping Favored	Heterogeneous Grouping Favored	Mixed Results	No Significant Differences
2-5	Smith (60)			
3,4				Stern (72)
4				Putbrese (71)
5				Eddelman (71)
5,6	Harbourger (64)			
6	Dewar (61)			
6				Wallen & Voules (60)
7				Campbell (64)

In the other variation students are regrouped at the beginning of each new topic, usually on the basis of a pretest on that topic. Studies of this variation are listed in Table VII. The results for this variation are also quite encouraging.

It is curious to note only one of these studies was done with elementary school students.

TABLE VII

Effects of Regrouping

Grade Level(s)	Homogeneous Grouping Favored	Heterogeneous Grouping Favored	Mixed Results	No Significant Differences
6	Lerch (60)			
7	Bierden (68)			
7	Stevenson (66)			
7				Willcut (67)
11		Mortlock (69)		
C	McBride (70)			
C				Merritt (72)

Have Criteria Other Than Ability Been used for Grouping?

Metfessel (1960) and Yerry (1964) both formed classes that cut across normal grade levels and thus were artifically heterogeneous with respect to age. Only slight differences in mathematics achievement were observed.

Savard (1960) and (1961) studies "limited range" grouping. His classes consisted of average children plus a small group of either above average or below average children. In general, students did better in mathematics achievement than corresponding students had done the previous year when heterogeneous grouping had been used.

Schmidt (1960) studied intra-class grouping on the basis of a sociometric test. The experimental students gained more than the control students did.

Steffani (1970) grouped calculus students on the basis of their college majors. This procedure produced slightly better results than the heterogeneous grouping that had been used the previous year.

Thelen (1967) grouped students on personality characteristics which were thought to be related to "teachability". Not surprisingly, the experimental students received higher grades than the control students. However, they did no better on standardized tests.

The Pygamalion Effect.

Schrank (1967) reports that, when students had actually been randomly assigned to classes but believed they had been grouped by ability, all classes performed at about the same level.

Tuckman and Bierman (1971) arbitrarily reassigned some low and middle ability students to the next higher level. This procedure did not increase achievement.

Does Ability Grouping Affect Student Attitudes?

In a substantial number of these studies listed in Part B some measures of student attitudes were taken. A wide variety of measures were used, including attitudes towards subject matter, toward self, toward school, towards the grouping procedure, etc.

These studies are listed in Table VIII. An inspection of this table shows that ability grouping in general has little effect on student attitudes or else changes them for the better.

TABLE VIII

Effects of Ability Grouping on Student Attitudes

Grade Level(s)	Attitudes Improved	Attitudes Declined	Mixed Results	No Significant Difference
1				Johnston (73)
2-3	Karnes et al (63)			
3,4	Stern (72)			
3-5				Hofset (70)
3-5				Hofset (71)
3-6		Barker Lunn (70)		
3-6				Ferri (71)
3-6		West & Sievers (60)		
4				Lerch (60)
4-11			Borg (64)	
5,6			Goldberg et al (66)	
6				Morgenstern (63)
6				Bachman (68)
7	Bierden (68)			
7	Willcut (67)			
7,8				Mikkelson (62)
7,8				Peterson (66)
8		Wilcox (63)		
9	Alam (68)			
10				Lovell (60)
11	Mortlock (69)			
12	McBride (70)			
12				Merritt (72)

Conclusions

Three recommendations emerge clearly from the above discussion.

First, school systems should be encouraged to experiment with homogeneous grouping for students who are high in mathematical ability, but to group the remaining students heterogeneously.

Second, teachers should be encouraged to experiment with within-class grouping, particularly the variation which uses regrouping at the beginning of each new topic.

A third recommendation is addressed to researchers in mathematics education. Research on team-teaching of mathematics is quite skimpy and not very encouraging about the effectiveness of such a teaching procedure. But many of the school buildings recently constructed are designed with open spaces and hence can accommodate team-teaching.

It would seem worthwhile to investigate the effectiveness of team-teaching using the procedures of the Dewar and Hombourger studies. With a team of three, say, teachers, one of them could teach the beginning of a chapter, or topic, to the combined group of students for two or three days, after which the students could be separated into a high, a middle, and a low group for the rest of that topic. Each teacher could then work with one of these groups for the rest of the chapter or topic.

The difference between this procedure and that used in the Dewar and Hambourger studies is that each student would be in direct contact with a teacher all the time, rather than only one third the time.

The results of within-class regrouping shown in Table VII suggest that this procedure might well be an effective way of making use of open space schools.

PART B

ANNOTATED BIBLIOGRAPHY

Adamson, David P. Differentiated Multi-Track Grouping vs. Uni-Track Educational Grouping in Mathematics. Doctoral dissertation, Brigham Young University (1971).

Three hundred sixty seventh and eighth grade students were used in this study.

These students were in two schools. In one of the schools homogeneous grouping was used. Students were assigned to one of three ability levels on the basis of academic achievement and teacher recommendations. Instruction was tailored to the ability level of the class. The middle ability level classes were provided with teacher aides. Fifteen boys and girls at each of these levels, in each of the two grades were selected for analysis.

In the control school heterogeneous grouping was used. The comparison sample was chosen on the basis of IQ.

Findings:

1. On a computation test the experimental students did better than the control students.
2. On arithmetic concepts and on arithmetic applications there were no significant differences.

Alam, Sami Jamil. A Comparative Study of Gifted Students Enrolled in Separate and Regular Curriculums. Doctoral dissertation, Wayne State University (1965).

Thirty two ninth grade students were included in this study. Sixteen of the students had been selected at the end of grade two for a program for gifted students. The selection was based on IQ. The matched control group had not been included in the program.

The experimental students were homogeneously grouped during the entire program. The emphasis was on activity and individualized work.

A standardized test was aministered at the end of the ninth grade. A self-concept scale was also administered, as was a critical thinking instrument.

Findings:

1. There were no significant differences between the experimental students and the control students on arithmetic computation or on arithmetic concepts.
2. The experimental students scored higher than the control students on arithmetic applications.
3. The experimental students were higher on critical thinking than the control students.
4. There was no difference between the experimental and control students in self-concept.
5. The experimental students were higher in leadership than the control students, but there were no other significant personality differences.

Eichman, Alfred M. Factors Related to the Achievement of Junior High School Students in Mathematics. Doctoral dissertation, University of Oregon (1968).

Three hundred twenty four seventh grade students were analyzed in this study. They came from 15 schools (grades 1-8) in a large city.

Two variables were investigated. Some of the schools were departmentalized for mathematics and some were not. Within each pattern some classes were grouped according to ability and some were not.

A mathematics posttest was administered at the end of the seventh grade. Measures of self-concept were also obtained at that time.

Findings:

1. There was no significant difference on mathematics achievement between the four patterns.
2. There were no significant differences on mathematics self-concept between the four patterns.

Bailey, Herman P. A Study of the Effectiveness of Ability Grouping on Success in First Year Algebra. Doctoral dissertation, St. Louis University (1967).

Eighteen ninth grade algebra classes were used in this study.

In the school system in which the study took place high school students were tracked for each subject on the basis of previous achievement in that subject. There were three tracks or ability levels.

Two heterogeneously grouped classes were formed by moving the even numbered students from a track one algebra class into a track two class and by moving the even numbered students from the track two class up to the track one class.

A standardized algebra test was administered at the beginning and at the end of the year.

In a follow-up study a standardized geometry test was administered to those students who had continued from algebra into geometry.

Findings:

1. On a standardized test the high ability students in the heterogeneously grouped class did as well as comparable students in the homogeneously grouped classes.
2. High ability students in heterogeneously grouped classes received higher grades from their teachers than comparable students in homogeneously grouped classes.
3. For medium ability students the grouping method had no effect on test scores or on teacher assigned grades.

Eglow, Bruce; Curtin, James. Ability Grouping of Bright Pupils. Elementary School Journal, Vol. 66 (1966) pp. 321-326.

Bright third grade students were studied. There were 150 of them included in the study.

These students were at a university laboratory school, most of whose students were of high ability. School records were sampled until 150 complete sets of data were located. The data included IQ and achievement scores.

Half the students were randomly assigned to a heterogeneous set and the remainder were sorted by IQ into three sets of 25 each. The IQ ranges for these were 100 to 122, to 141, and 142, to 181.

The variances, for all achievement scores, were computed for each of the four sets.

Findings:

1. On each of arithmetic computation, arithmetic knowledge, and arithmetic problem solving, the variances were not significantly different.

Balow, Irving H. The Effects of "Homogeneous" Grouping in Seventh Grade Arithmetic. The Arithmetic Teacher, Vol. 11 (1964) pp. 186-191.

Seventh grade students were studied. They totaled 353.

When they entered the seventh grade they were placed in homogeneous fast, average, or slow sections on the basis of achievement test scores at the end of grade 6, letter grades in arithmetic, and teacher comments.

A computation test and a problem solving test were administered at the end of the second week of school. The same tests were administered again at the end of the school year.

Findings:

1. More than half, 186, of the students had initial computation scores between the lowest score of the fast sections and the highest score of the slow sections and so any of the three levels would have been appropriate. For problem solving, the overlap was greater. Two hundred ninety-nine could have been placed at any of the three levels. On a composite of the two scores, the overlap was 255 students.
2. The overlap at the end of the year was 231 for computation and 230 for problem solving.
3. Over the year, the fast sections gained the most on computation and the slow sections the least.
4. In problem solving, correlations between gain and initial score were all negative, but there was heterogeneity of regression. The negative correlation was greatest for the slow sections.

Balow, Irving H.; Ruddell, Arden K. The Effects of Three Types of Grouping on Achievement. California Journal of Educational Research, Vol. 14, (1963) pp. 108-117.

Sixth grade students in four schools were studied. About 200 students had complete records and could be included in the analysis.

In one school, students were grouped homogeneously in four mathematics ability levels, and similarly for reading. For all but these two classes, they were grouped heterogeneously.

In a second school, the two sixth grade classes were homogeneously grouped on the basis of IQ.

In the other two schools, heterogeneous grouping was used exclusively.

Standardized tests were given at the beginning and the end of the school year. IQ information was obtained, and three IQ levels were used in the analysis.

Findings:

1. The IQ-grouped students gained more than homogeneously grouped students in arithmetic computation, with the heterogeneously grouped students making the least gain.
2. There was no significant treatment difference on arithmetic problem solving (or on other subjects, such as social studies, science, spelling, etc.).
3. Although the middle IQ groups showed greater gains than either of the other two on most subjects, the homogeneously grouped high IQ students gained more than the middle or low groups.

Barker-Lunn, Joan C. Streaming in the Primary School. National Foundation for Educational Research in England and Wales (1970).

The students in this study were followed over a four year period from the beginning to the end of junior school (ages 7+ to 11+). There were fifty schools which practiced streaming and fifty matched schools which did not stream. The total number of students was over 7,000 to begin with. Attrition resulted in 36 matched pairs of schools at the end of the study including over 5,000 students.

A mechanical arithmetic test and a problem arithmetic test were administered at the end of each year. However, there were two parallel forms. A random half (set A) of the pairs used one form at the end of the first, second, and fourth years and the other form at the end of the third year. The remaining pairs of schools (set B) used each year the form other than the one used by set A. (Thus there were really two parallel experiments.)

In addition to the above tests a number concept test was administered at the end of years two, three, and four.

A number of attitude tests were also administered.

Students were divided into three ability levels on the basis of a reading test at the beginning of the study. Students were also divided into three SES levels. The results were analyzed separately by sex using ANCOVA. Thus there were two × three × three × four = 72 comparisons made for the mechanical arithmetic and problem arithmetic tests. For concept arithmetic tests there were 54 comparisons.

Findings:

1. For set A the number of significant differences were as follows:

	Unstreamed	Streamed
Mechanical arithmetic	13	1
Problem arithmetic	10	1
Number concept	4	1

For smaple B significant differences were as follows:

	Unstreamed	Streamed
Mechanical arithmetic	1	15
Problem arithmetic	1	13
Number concept	1	2

2. The discrepancies between the results of the two different experiments seems to be entirely due to the sampling. In particular, nothing known about the teachers explained this discrepancy.

(The overall conclusion is that mathematics achievement is about the same in streamed and in unstreamed schools.)

5. However, for self-concept, attitudes toward school, etc., there were no significant differences for high ability students, but in general the results were more positive for the unstreamed schools than for the streamed schools.

Becker, Leonard J. An Analysis of the Science and Mathematics Achievement of Gifted Sixth Grade Children Enrolled in Segregated, Partially Segregated and Non-Segregated Classes Doctoral dissertation, Syracuse University (1963).

The sample for this study consisted of 235 sixth grade students in 26 different classes. All the students were considered "gifted" in that each had an IQ of 130 or more.

Of these students, 67 were in heterogeneously grouped classes. Another 70 were in special programs for gifted students in which they were grouped separately for part of the school day or week for enrichment purposes. The remainder of the students were in homogeneously grouped classes.

A common IQ test was administered to all the students and a standardized test was administered at the end of the school year.

Findings:

1. The mean mathematics scores for the three programs were not significantly different. (This was not true for science scores.)
2. Boys scored higher than girls.
3. There was an interaction between program and sex. Boys in the special programs scored higher than the boys in the other kinds of classes. This was not true for girls.

Bierden, James E. Provisions for Individual Differences in Seventh Grade Mathematics Based on Grouping and Behavioral Objectives: An Exploratory Study. Doctoral dissertation, University of Michigan (1968).

Forty-four seventh graders were used in this study.

Behavioral objectives drawn from the textbook were written by the investigator at three levels: basic, intermediate and advanced.

The objectives for each topic were distributed when the topic was first taken up. The investigator then taught the topic to the class as a whole, aiming the instruction at the intermediate level of objectives.

This phase was followed by a test on the basis of which each student was assigned to one of three ability groups.

During the next phase each group worked separately and then a test appropriate to the ability level of the group was administered.

During the final phase the students worked individually.

During the course of the study some changes were made in the design. In particular the final phase was omitted for some topics.

A variety of mathematics and attitude tests was administered.

This was primarily an exploratory study. In particular, there was no control group.

Findings:
1. Students did not end up at the same level for each topic.
2. Student attitudes to these procedures was favorable.
3. Students did learn a significant amount of mathematics during the school year.

Borg, Walter R. An Evaluation of Ability Grouping. Cooperative Research Project No. 577, Utah State University, Logan, Utah (1964).

Five sets of students were included in this study. Sample IV was in grade four at the beginning of the study, Sample VI in grade six, Sample VII in grade seven, Sample VIII in grade eight, and Sample IX in grade 9. Approximately 2,000 were involved during the first year. Another 2,000 were added at the beginning of the second year.

The classes which made up these samples were chosen in two similar school districts. In district A, classes were grouped by ability on three levels in the elementary school. The grouping, which was done each year, was made on the basis of achievement scores from the preceding year. At the secondary level, grouping for mathematics was based on mathematics achievement the previous year. No grouping was needed for the advanced courses.

In this district, the rate of presentation of the subject matter was adjusted to the ability level of the class. Curricular materials were also adjusted to the ability level, but the main difference between classroom activities at different ability levels was speed at which material was presented.

In district R, random grouping was used together with enrichment to adjust the level and depth (rather than the rate) of presentation of material.

A standardized test was administered at the beginning and end of the first year. After that, another standardized test was used for each of the final three years.

Findings:
1. The only significant differences were as indicated in the tables:

Sample IV.	Superior	Average	Slow
Year 1	A > R		
Year 2		R > A	
Year 3			
Year 4	A > R		

Sample VI.	Superior	Average	Slow
Year 1	A > R	A > R	
Year 2	R > A		R > A
3 & 4			
all 4	R > A		

Sample VII.	Superior	Average	Slow
Year 1		R > A	R > A
Year 2	A > R		
3 & 4			
all 4			

Sample VIII.	Superior	Average	Slow
Year 1		R > A	
Year 2	A > R		A > R
3 & 4			
all 4			

Sample IX.	Superior	Average	Slow
Year 1			
Year 2			
3 & 4	A > R	A > R	
all 4		A > R	A > R

2. On one measure of attitudes toward school, in Sample VI the superior and average students in District R scored higher than those in District A. In sample VII this was true for average students.
3. On another attitude battery, Samples IV, VI, VII, and VIII combined, there were no significant differences between the districts on attitudes toward peers. Both the superior boys and girls and the slow boys in District A had better attitudes towards their teachers than the corresponding students in District R. Slow boys in District A had a better attitude toward school than slow boys in District R. No other comparisons showed a significant difference.
4. There was very little difference between the two systems on a test of Ideal Self Concept. For Actual Self Concept, the results favored District R. The girls were largely responsible.
5. There were no significant differences with respect to test anxiety.

Borg, Walter R. Ability Grouping in the Public Schools. Journal of Experimental Education, Vol. 34 (1965) pp. 1-97.

A slightly edited version of the above report.

Bouri, Janet; Barker Lunn, Joan. Too Small to Stream. National Foundation for Educational Research in England and Wales, Occasional Publication Series, No. 21 (1969).

Fourteen matched pairs of small (fewer than eight classes) junior schools were studied.

In one of each pair students were assigned to classes by age which resulted in heterogeneous ability grouping. In the other school in each pair students were retained or promoted in such a way as to create classes

somewhat homogeneous in ability. Tests in problem arithmetic, mechanical arithmetic, and concept arithmetic were given to all students in each school.

Findings:
1. On problem arithmetic there were four significant differences in favor of heterogeneous grouping. There were twenty eight non-significant differences in favor of heterogeneous grouping and four non-significant differences in favor of homogeneous grouping.
2. On the second test there 18 non-significant differences in favor of each form of grouping.
3. On the third (not taken by the seven year olds) there were 13 non-significant differences in favor of heterogeneous grouping and 14 in favor of homogeneous grouping.
(In English and in reading there were pronounced advantages for homogeneous grouping.)

Campbell, Azzie L. A Comparison of the Effectiveness of Two Methods of Class Organization for the Teaching of Arithmetic in Junior High School. Doctoral dissertation, Pennsylvania State University (1964).

This study utilized 240 seventh grade students.

The students were in two junior high schools. They had all been tested for arithmetic achievement at the end of the sixth grade. A total of 220 pupils in each junior high school were given IQ and arithmetic achievement tests at the beginning of the seventh grade.

On the basis of these scores, four classes of 30 each were formed in each school in such a way that they were all matched in IQ and in arithmetic achievement. The classes in each school were randomly assigned to the experimental treatment.

Each instructor involved taught one experimental and one control class.

The experimental treatment was within-class grouping. Each experimental class was divided into a high, an average, and a low group on the basis of the achievement test given at the beginning of the year and also on the teacher's judgement, based on teacher-made tests and daily class performance.

The subgroups were located in different parts of the room. The teacher spent equal amounts of time with each ability group. The textbook level was adjusted to the ability of the group.

The control classes were taught in the usual whole-class method.

Another form of the standardized achievement tests was given at the end of the school year.

Findings:
1. The original matching of the classes on IQ and achievement was successful.
2. Both experimental and control students made significant gains in achievement during the year.
3. There was no significant difference between the gains of the experimental and the control students.

Campbell, L. M.; Knill, W. D. Ability Grouping and Grade Nine Achievement. Alberta Journal of Educational Research, Vol. 11 (1965) pp. 226-232.

The subjects of this study were all the ninth grade students in a single school over an eight year period.

During the first three years, heterogeneous grouping was used. During the next three years, ability grouping was used, based on academic achievement in the eighth grade. During the last two years, grouping was done by subject matter area, based on eighth grade achievement in that subject.

Subject matter stanines and ability percentiles were collected.

Thirty-two matched pairs at each of three IQ levels were selected for each of the three time periods. Pairing was done on the basis of sex, age, ability, and IQ.

Also, for a second analysis, a similar number of cases was chosen randomly.

Findings:

1. A comparison of the first two 3-year periods found that homogeneous grouping was significantly better than random grouping for the top and bottom ability levels. There was no significant difference for the middle ability group. This result held for both sampling methods. (The results for other subject matter areas were different.)
2. When subject matter grouping was compared with heterogeneous grouping, the results were the same as in the comparison between heterogeneous and homogeneous grouping.
3. Subject matter grouping was no more effective than homogeneous ability grouping.

Chiotti, Joseph Frank. A Progress Comparison of Ninth Grade Students in Mathematics from Three School Districts in the State of Washington with Varied Methods of Grouping. Doctoral dissertation, Colorado State College (1961).

This study involved 62 ninth graders in each of three junior high schools.

In one of these schools students were grouped on the basis of past achievement (and teacher recommendations when in doubt) at five different levels. Each level cut across the three grades in the school. Assignments to levels were reviewed each semester.

In the second school three ability levels were defined within each of the three grades and students were grouped homogeneously by ability.

In the third school heterogeneous grouping was used.

A standardized mathematics test was administered at the beginning and at the end of the school year.

Pretest scores were used to establish comparable samples within each school for comparison purposes.

Findings:

1. The grouping patterns had no effect on achievement.

Cluff, James E. The Effect of Experimentation and Class Reorganization on the Scholastic Achievement of Selected Gifted Sixth Grade Pupils in Wichita, Kansas. Doctoral dissertation, University of Arkansas (1964).

There were forty eight experimental and forty eight control students in a two year study covering grades four and five.

The experimental students were chosen from among those with an IQ of 120 or higher and with a grade equivalent of at least 4 1/2 years. The control students were matched with the experimental students on sex, age, IQ, and achievement.

The experimental students were homogeneously grouped in four classes (there were eighty experimental students to start with, but because of attrition and the need to equalize cells in an analysis of variance only 48 pairs could be included in the analysis). The control students were grouped heterogeneously.

The experimental students were not accelerated but rather were urged to broaden their understanding in all of their classes.

A standardized achievement test was used at the beginning of the fourth, fifth, and sixth grades.

Findings:

1. On arithmetic skills the experimental students started 1/2 of a month behind the control students and were 2 1/2 months ahead at the end of 2 years.

Dahllof, Urban S. Ability Grouping Content Validity, and Curriculum Process Analysis. Teachers College Press, Teachers College, Columbia University, New York, New York.

Dahllof used the same students as were included in Svensson's study.

Dahllof had collected various kinds of information about the schools in which the Svensson study took place. He used this information in reanalyzing Svensson's data.

Findings:

1. The students who were placed in the high ability group at the end of the fourth grade had about the same mathematics score at the end of grade six as the students who were in homogeneously grouped classes during grades five and six. However, the selected students spent much less time on the material covered by the sixth grade test.
2. In grades seven and eight the students who had been selected at the end of the fourth grade spent much less time on arithmetic and more time on algebra and geometry than the students selected at the end of sixth grade. However, the test used at the end of grade seven was largely on arithmetic and hence did not test much of the material learned by the early select students. To a lesser extent the same thing was true for the eighth grade students.

Daniels, J. C The Effects of Streaming in the Primary School. II. -- A Comparison of Streamed and Unstreamed Schools. British Journal of Educational Psychology, Vol. 31, (1961) pp. 69-78.

Four junior schools, each of a size to admit three classes each year, furnished the subjects for this study.

Schools A and C were under one LEA and schools B and D in another. In most respects, each pair of schools was quite similar except for philosophy about streaming. Schools A and B were against and did not use it while schools C and D were for it and did it.

Samples from schools A and C were chosen that were matched on IQ. Since test results were available for three consequetive year-groups, there were three such matched samples.

(Although it is not so stated, evidently the same procedure was used with schools B and D.)

Findings:

1. School A and School C were the same on IQ during the first year, but School A was higher at the end.
2. On each of six comparisons, over the four year period, School A scored higher than School C in arithmetic.
3. School B was consistently ahead of School D on IQ.
4. At the end of four years, school B had gained more in arithmetic than School D.

Harris, O. L. Jr., Tracy, Neal H. Arithmetic Achievement and Instructional Grouping. The Arithmetic Teacher, Vol. 10 (1963) pp. 12-17.

This study involved 393 students in grades 4, 5, and 6 in two schools.

In one of the schools, students were grouped for arithmetic on the basis of past achievement and ability. No grouping was done across grade levels. In the other school, random assignment to classes was used.

A standardized arithmetic test, an ability test, a self-concept scale, two anxiety scales, and an arithmetic attitude inventory were administered during the third week of school. The arithmetic achievement test was administered again at the end of the school year.

Analysis of covariance used all the initial achievement, cognitive and affective scores as covariates.

Findings:

1. At the beginning of the school year, the school using ability grouping scored higher on achievement. At the end of the year this advantage had vanished. The randomly grouped students gained significantly more on both computation and reasoning.

Dewar, John A. An Experiment in Intra-Class Grouping for Arithmetic Instruction in the Sixth Grade. Doctoral dissertation, University of Kansas (1961).

Eight classes of sixth grade students provided the subjects for this study.

Eight teachers volunteered to take part in the experiment. They were randomly assigned to the experimental (within-class grouping) treatment or to the control (whole-class procedure) treatment.

At the beginning of the school year, a standardized arithmetic achievement test was administered and also an IQ test. Using the achievement information, (but not the IQ) and previous school records, each experimental class was divided into three ability levels. The initial grouping was reviewed and, when appropriate, revised after a ten day period. After that, no revisions in the grouping were made. A separate curriculum outline was provided for each level.

Teachers divided their time equally between their three experimental groups.

The control teachers followed the normal whole-class teaching procedure.

An attitude inventory was administered at the end of the school year, and an alternate form of the achievement test.

Findings:

1. The experimental and control groups were comparable at the beginning of the study.
2. The experimental group made greater gains than the control group.
3. The above and below average experimental groups significantly exceeded their controls, but the average students did not.
4. Both the experimental teachers and the experimental students approved.

Dewar, John A. Grouping for Arithmetic Instruction in Sixth Grade. Elementary School Journal, Vol. 63 (1963) pp. 266-269.

A brief account of his dissertation findings.

Echternacht, Charles; Gordon, Virginia. Breaking the Lock Step in Arithmetic. The Arithmetic Teacher, Vol. 9 (1962) pp. 86-89.

This study involved one hundred fifty students in grades four, five, and six.

On the basis of IQ, previous mathematics achievement, and teacher recommendations, the students were placed in five different classes, each as homogeneous as possible. The classes cut across the three grade levels.

The grouping was flexible and students could be moved up or down at any time.

The teachers grouped within classes, usually organizing two subsets.

Findings:

1. The lowest group completed all the work normally done in grade four. Each of the other groups completed more than a normal years work.
2. On a standardized test administered in the middle of the year in both the first and the second year of the study, gains (for classes, not students) ranged from 1.2 to 2.2 years.

Eddelman, Virginia K. A Comparison of the Effectiveness of Two Methods of Class Organization for Arithmetic Instruction in Grade Five. Doctoral dissertation, Northeast Louisiana University (1971).

One hundred fifty six fifth graders in six classes were involved in this study.

In a large school system one elementary school was randomly chosen at each of three SES levels. Within each school two fifth grade classes were chosen at random, one to be an experimental class and the other a control class.

In each experimental class the students were divided into three ability levels. The decisions were based both on teacher opinions and the results of a standardized mathematics pretest. Teachers spent equal amounts of time each day with each of the subsets.

In the control classes (taught by the same teachers) the students were separated into three subsets by a random process. The procedure in each of these control classes was otherwise the same as in the experimental classes.

The treatment lasted for nine weeks after which a parallel form of the standardized test was administered.

Findings:
1. There was no significant treatment effect.
2. Teacher opinions about intra-class grouping (obtained before the data were analyzed) were not in accord with the above finding.

Enzmann, Arthur M. An Evaluation of the Science and Arts Curriculum for Selected Students of High Ability at Cass Technical High School, Detroit, Michigan. Doctoral dissertation, Wayne State University (1961).

Several hundred gifted eighth graders were invited to participate in a special program in a technical high school in a large city. At the end of grade twelve, two hundred six students had complete scores, ninety four of whom had participated in the special program and a hundred twelve who had declined the invitation.

In mathematics the special program was acceleration.

Standardized criterion tests were administered at the ends of grades ten and twelve.

Pairs of students matched on IQ, sex, and mathematics scores at the end of grade eight were compared.

Findings:
1. The only significant difference was that the students in the accelerated program did better than the control students on a standardized mathematics test at the end of grade ten. (However, the standardized test used at the end of grade twelve did not cover much of the material studied by the accelerated students).

Ferri, Elsa. Streaming: Two Years Later. National Foundation for Educational Research in England and Wales (1971).

The students in this study, slightly over 1700 in all, constituted the bulk of those that came from 14 matched pairs of schools that had been included in the Barker Lunn study. The students were now at the end of their second year of secondary school.

Placement procedures in the secondary schools were rather complex but arbitrary cutting point was set and the schools, secondary modern comprehensive, and grammar, were classified as being either streamed or not streamed.

Two mathematics tests were administered at the end of their second year. In addition, attitude and personality scales were administered.

For the analyses, the students were stratified on ability and SES.

Findings:
1. On mathematical concepts, out of nine comparisons for boys, there were two cases where the non-streamed (in junior schools) students were significantly better.

3. On mathematics problems, out of nine comparisons the non-streamed (in junior schools) students excelled twice and the streamed students twice.
4. For the girls there were no significant differences due to streaming in junior school.
5. For boys streamed in secondary school, the non-streamed students exceeded the streamed students in one out of five comparisons both in mathematical concepts and in mathematical problems.
5. For girls, streaming in the secondary school had no effect.
6. On attitudes, the results were generally inconclusive. However, high ability students tended to go down while low ability students, especially from non-streamed junior schools, went up when they were in non-streamed secondary schools.

French, John W. Evidence from School Records on the Effectiveness of Ability Grouping. Journal of Educational Research, Vol. 54, (1960) pp. 83-91.

In this study, more than 700 tenth grade geometry students, in two different years, were studied in one large high school.

An IQ test had been administered to all students. A uniform final examination was used for all classes.

The standard deviation on the IQ test was computed for each class and used as measure of heterogeneity of the class.

The regression of class achievement means on IQ was computed and used to measure the effectiveness of each teacher.

Ability quartiles were defined for each class using IQ. Also classes were classified as high or low ability on the basis of IQ.

Findings:
1. Teacher effectiveness was unrelated to heterogeneity of the class.
2. Both high and low ability students achieved better in high ability classes than in low ability classes, but the difference was not significant.

Frye, Charles H. Group vs. Individual Pacing in Programed Instruction. Oregon State System of Higher Education. Title VII Project Number 847, National Defense Education Act of 1962.

Forty four high school freshmen were in this study.

The students were taking the beginning algebra course. They could solve quadratic equations by factoring and by completing the square, but they did not know the quadratic formula.

The students divided into two sets consisting of pairs matched on a standardized ability test and on an algebra prognosis test. From the top and bottom quartiles of the ability distribution, two separate sets were selected in such a way that each student in one set was matched on the basis of algebra ability with a student in the other set. Half the students in each set came from the top quartile and half from the bottom quartile.

Two homogeneous sets were established in the same way using the students in the second and third quartiles of the distribution.

One of the two heterogeneous sets (HET GP) went through a program on the quadratic formula as a group. The program was administered by means of a slide projector and each frame was presented only after all members of the set had responded to the previous frame. The other

heterogeneous set (HET IP) went through the same program in the standard individual manner.

One of the homogeneous sets (HON GP) went through the program as a group and the other (HOM IP) went through the program individually.

A criterion test on the quadratic formula was administered when the students finished the program. If a student did not receive a perfect score he was asked to review that part of the program which he was having trouble with and then take a parallel form of the criterion test. (Only two students had to review more than once.)

The total time inclusing review time to complete the program was recorded for each student.

Findings:

1. Time for HET GP was greater than time for HET IP.
2. Time for HOM GP was not significantly different from HOM IP.
3. Time for HET GP was greater than time for HOM GP.

Gilbert, Virginia; McKie, Earl. Skill-Level Grouping in Modern Mathematics K-6; Final Report, Clark County School District, Las Vegas, Nevada. ED 033 047

The students involved in this study were all the second through sixth graders in two elementary schools.

The schools were in a disadvantaged area. One school contained an experimental program and the other school was the control. On the basis of standardized tests the students in the experimental school in grades two and three were placed in seven Skill-Level Groups. Students in grades 4 through 6 were placed into six Skill-Level Groups.

The grouping procedure was flexible. Students could be moved up or down at any time.

Instructional aids were provided for the experimental program. A mathematics specialist worked with the teachers.

In the control school no grouping was used and the normal instructional program was followed.

Standardized achievement tests were administered at the end of the year and a special mathematical concepts test was also used in grades 5 and 6.

The program lasted 31 weeks for grades 4 through 6 and 26 weeks for the second and third graders.

Findings:

1. For grades 4 through 6 the only significant differences were that the sixth grade experimental students did better than the sixth grade control students on the mathematical concepts test and at the fourth grade level the control students did better than the experimental students on the standardized posttest.
2. In both grades two and three the control students were scored higher than the experimental students both in the pre-test and on the posttest.
3. The greatest amount of growth occurred among those farthest below grade level in grades two, three, and four.

Goldberg, Miriam L.; Passow, A. Harry; Justman, Joseph. The Effects of Ability Grouping. Teachers College Press, Teachers College, Columbia University, New York (1966).

The students involved in this study were in grades 5 and 6. Approximately 2200 students remained in the study over the two year period and had complete pre- and posttests scores.

Five ability levels were specified labeled from A (IQ greater than or equal to 130) to E (IQ < 100). Classes were established that represented all patterns of ability levels that satisfied the condition that when two different ability levels were represented in the class, so were all the intermediate levels. These classes were kept intact over the two year period covering grades 5 and 6. However, there was no restriction on the assignment of teachers to classes, instruction to teachers, etc. Normal school procedures were followed.

A standardized achievement test and attitude test were administered at the beginning of the fifth grade and at the end of sixth grade.

Findings:

1. Using the composite arithmetic score as a criterion measure, it made no difference to students not in the top level whether or not top level students were included in the class.
2. Students in classes containing all three middle levels did better with A level students, but not E level, also in the class than with Es but without As.
3. For bottom level students it was better to have A level students in their class than not to have them.
4. For other levels, it was better not to have E level students in the class.

(The pattern of results for other subject matter areas was quite different.)

5. For all ability levels taken together, broad range classes (at least four ability levels) were more effective than narrow range classes (no more than 2 ability levels) for arithmetic computation, arithmetic concepts, and arithmetic reasoning. In addition, for both reasoning and computation a broad range class was more advantageous than a medium range class (3 ability levels) for arithmetic reasoning and arithmetic computation. However, there was no significant difference between medium range and narrow range classes on any of these three criteria.
6. For the individual ability levels, of the 45 possible comparisons only 5 were significant. For level B a broad range was more advantageous than a medium range for arithmetic computation. For level C a broad range was more advantageous than a narrow range for arithmetic reasoning and arithmetic computation and more advantageous than a medium range for arithmetic computation. For level D a broad range was more advantageous than a medium range for arithmetic computation.
7. For arithmetic computation there was positive correlation between ability and achievement increment. However, for concepts the correlation was negative. However, this might be due to a ceiling effect for the standardized test.
8. For arithmetic computation it was better to be in a class with just one ability level or in a broad range class than to be in a medium range class, whether at the top, the middle, or the bottom. Otherwise there was no significant differences for relative position in class.
9. It made no difference on any arithmetic score to be in the top or the bottom level of a two or three level class.
10. Variability in achievement was greater over classes than over ability for all grouping patterns.

11. Correlations between achievement and attitudes towards school were generally positive but small.
12. Teachers were more successful in handling several ability levels for one subject than in handling several subjects for a single ability level.
13. Over all subject matter areas, self-concept was better in narrow range classes than in broad range classes for the less able students.
14. Actual self concept (not mathematics specific) scores of the C and D level students were enhanced by the presense of gifted students. For E level students, the presense of A level students depressed this score. The A level students had no effect on the scores of B level students.
15. Ideal self concept scores were higher whenever the A level students were not included.

Goldberg, Miriam L.; Passow, A. Harry. The Effects of Ability Grouping. Education, Vol. 82 (1962) pp. 482-487.

A brief review of the above report.

Hambourger, James M. Grouping for Arithmetic Increases Pupil Growth. Chicago Schools Journal, Vol. 45 (1964) pp. 362-365.

One class of fifth and sixth graders was studied, thirty three students in all.

Students were grouped within the class at three levels.

They were tested at the end of each week. Three versions of each test were used. All students tried the intermediate level test and then went on to the one best suited to their abilities.

Each student received twenty minutes of direct teaching each day and also had another twenty minutes of seat work.

Findings:
1. The sixth grade students gained 1.6 years on a standardized mathematics test.
2. Only 9 students had been at grade level or above at the beginning of the year. By the end of the year there were 18.

Hofset, Arnold. Achievement and Attitude towards School in Classes with Many and with Few Low Ability Pupils. Scandinavian Journal of Educational Research, Vol. 14 (1970) pp. 39-51.

This study included all the third and fifth grade students in a Norwegian town. The numbers were 687 and 607. They were also tested when they reached grade 7.

Using stanines in mental age, three ability levels were specified. The top three stanines were placed in H, the middle stanine (5) in M and the bottom three stanines in L. Students in stanines 4 or 6 were excluded from the anlysis.

Plus-classes contained four or more with stanine ability scores of 8 or 9. Minus-classes had no more than one such student.

A standard arithmetic test was aministered at the end of the school year to the third and fifth grade classes.

An attitude inventory was also administered at the end of the year.

Findings:
1. Class type had no significant effect on achievement.
2. There were no interactions.
3. Class type had no effect on attitudes.
4. Sociometric (leadership) status was lower, on the average, in plus-classes than in minus-classes. (The gifted students attract more of the leadership choices).

Hofset, Arnold. Gifted Pupils as Sources of Inspiration in the Class? Scandinavian Journal of Educational Research, Vol. 15 (1971).

The same students were used as in his 1970 report.

High, middle, and low ability were defined as in the other study.

Classes with fewer than 6 L students were labeled FB and those with 6 or more MB.

Chi-square tests showed that the distribution of stanines in the H-MB and H-FB classes were similar and similarly for the L classes.

Findings:
1. At no grade level was there a significant difference between corresponding FB and MB classes
2. There were no significant interactions between class type, ability, and sex.
3. Class type had no effect on attitudes.

Johnston, Herbert J. The Effect of Grouping Patterns on First-Grade Children's Academic Achievement and Personal and Social Development. Doctoral dissertation, University of Miami (1973).

The experimental students in this study constituted one pod of 99 first graders in an open space school. An additional 30 students in another pod constituted the control group.

The treatment covered the second half of the school year. The experimental students were grouped for arithmetic on the basis of a test. They were stratified into three heterogeneous groups. The control students were grouped homogeneously on three reading ability levels.

Pre- and post standardized mathematics tests were administered and a self concept inventory was also administered.

Findings:
1. There were no significant differences due to the two grouping procedures on either mathematics achievement or self concept.

Karnes, Merle B., et al. The Efficacy of Two Organizational Plans for Underachieving Intellectually Gifted Children. Exceptional Children, Vol. 29 (1963) pp. 438-446.

The subjects of this study were drawn from all the students in two large elementary schools serving an upper middle class area. A total of 48 was ultimately involved, in grades two through five. The study lasted three years.

Those with an IQ of 120 or higher had discrepancy scores computed between expected and actual achievement scores in mathematics and reading. Those more than one sigma below the mean on these discrepancy scores were designated as underachievers.

In one school, gifted students were grouped homogeneously, in the other heterogeneously.

Achievement and creativity tests were administered, and measures of creativity and peer acceptance were taken.

Findings:

1. The gifted underachievers in the homogenous classes had better academic achievement than those in heterogeneous classes.
2. The homogeneously grouped gifted underachievers did slightly better than those heterogeneously grouped in gains in creativity and preceived parent attitudes, but not in perceived peer acceptance.

Koontz, William F. A Study of Achievement as a Function of Homogeneous Grouping. Journal of Experimental Education, Vol. 30 (1961) pp. 249-253.

Five fourth-grade classes in one school provided the experimental students for this study. The control students were in other schools.

An achievement test had been administered the previous spring. Results were used to section the students into five levels of arithmetic, each level being put in a separate class. (Other classes were formed independently for language and reading, so a semi-departmentalized procedure was used.

The arithmetic texts were for fifth grade for the top level, fourth grade for the next two levels, and third grade for the bottom two levels.

Control students were matched with the experimental students on ability but were instructed heterogeneously.

Achievement was measured at the end of the year with a standardized test.

Findings:

1. The control group, at each level, had higher achievement than the experimental group.
2. There was no interaction between treatment and ability level.

Lerch, Harold H. A Study Concerning the Adjustment of Arithmetic Instruction to Certain Individual Differences. Doctoral dissertation, University of Illinois (1960).

Four fourth grade classes were used in this study, two experimental and two control.

In the experimental classes at the beginning of any new topic the objectives were listed and a pretest was given. On the basis of this test the students were grouped into two or three ability levels. The teacher worked separately with each group and the students were regrouped at the beginning of the next topic.

The study lasted three months.

Standardized achievement tests and an attitude inventory were given at the beginning and at the end of the study.

Findings:
1. The experimental students gained significantly more than the control students.
2. There were no significant differential attitude changes.

Lerch, H. H. Arithmetic Instruction Changes Pupils' Attitudes Toward Arithmetic. The Arithmetic Teacher, Vol. 8 (1961) pp. 117-119.

A review of earlier studies and a brief report of his dissertation findings.

Lovell, John T. The Bay High School Experiment. Educational Leadership, Vol. 17 (1960) pp. 383-387.

Approximately 250 high school sophomores were used in this study.

The students were ranked by ability. The odd-numbered one were given the experimental treatment. The first thirty formed one class, the next thirty a second class, etc.

The even-numbered students were grouped heterogeneously, and formed the control group.

Each teacher involved in the study taught both an experimental and a control class.

Standardized achievement tests were administered at the beginning and the end of the year. Some attitude measures were also taken.

Findings:
1. There was no significant difference between gains in algebra scores for the two treatments. In English, the experimentals outgained the controls.
2. In the algebra classes, no significant attitude differences were found.

Mahler, Fred L. A Study of Achievement Differences in Selected Junior High School Gifted Students Heterogeneously or Homogeneously Grouped. Doctoral dissertation, University of Houston (1961).

Six seventh grade and two eighth grade classes were used in this study.

Four of the eight classes served as controls.

Ability was defined in terms of both IQ and achievement. Three of the seventh grade classes and one eighth grade class were homogeneously high in ability. The remaining high ability students were assigned to heterogeneously grouped classes.

Standardized achievement tests had been administered at the end of the previous year and were administered at the end of the experimental year.

The high ability students in the homogeneously grouped classes were compared with comparable students in the heterogeneously grouped classes. A total of 25 pairs of students were compared.

Findings:
1. There was no significant difference in the arithmetic achievement of high ability students in homogeneously grouped classes as compared with those in heterogeneously grouped classes.
(On reading, the homogeneously grouped students did better than the heterogeneously grouped students.)

McBride, Ralph B. Flexible Grouping and Differential Instruction Based on Achievement of Behavioral Objectives in a Mathematics Course for Prospective Teachers. Doctoral dissertation, University of Michigan (1970).

This study involved eight eight college students in a mathematics course for prospective elementary school teachers.

Behavioral objectives were written for each of the five units in the course. These were written at only one level but some enrichment objectives were also prepared for the better students.

A list of objectives was provided to the students at the beginning of each unit. The investigator then taught the whole group. At the end of phase one a test was given and on the basis of the test results the students were grouped at three levels: basic, intermediate and enrichment.

The students had been required to keep free the hour following the lecture so that, with the aid of one teaching assistant, four groups could be accommodated during phase two.

At the end of phase two a test at the appropriate level was administered.

Grades depended on which test had been taken (there was some flexibility) and on scores on the tests.

Findings:
1. Grouping level varied from unit to unit (and so could not have been predetermined) but about fifty percent of the variance in the grouping levels was predictable.
2. There was a low correlation between achievement and ability.
3. Attitudes towards mathematics did not change significantly.
4. Test anxiety was reduced.
5. In comparison with the course the previous year and with a norming group for a test on mathematics for prospective elementary school students, this group of students did satisfactory work.

McLaughlin, Jack W. A New Approach to an Old Problem. The Arithmetic Teacher, Vol. 9 (1961) pp. 112-116.

All the fourth, fifth, and sixth grade students in one school took part in this study.

Students were grouped homogeneously, across grades, on the basis of previous arithmetic achievement and teacher recommendations.

It was possible for students to move up or down at any time.

All groups were required to spend at least one day a week on drill, but otherwise each group progressed at its own pace.

Findings:
1. On a standardized test, there was an average gain of three days achievement for each day of instruction.
2. Over a 29 day span, 64 students made more than a year's gain as compared to 37 who did not gain significantly.

Merritt, Paul W. The Effects of Variation in Instruction and Final Evaluations Procedures on Community College Beginning Algebra Classes. Doctoral dissertation, University of Michigan (1972).

The students in this study comprised 8 classes of beginning algebra in a community college. 161 of them ended up with complete data sets.

Two variables were manipulated. One of these was within class grouping. For each unit of the course approximately one third of the alloted time was used for whole class instruction. This was followed by a diagnostic test on the basis of which the students were grouped at three ability levels. For the remainder of the time devoted to that unit the instructor spent most of his time with the low group. The middle group worked either individually or collaboratively. The top group worked mainly individually and enrichment materials were available.

The other variable was the availability of the opportunity of taking a retest if the student felt that he had not done well enough on a unit test.

Two instructors each taught one of each of the four combinations of treatments.

Instructional objectives were provided for all students including a sample test for each unit. Exercise sheets at three difficulty levels were also provided for each unit.

The study lasted one semester.

A standardized algebra test was used at the beginning and the end of the semester and attitude inventory was also administered at the beginning and end as was a test anxiety scale.

Findings:

1. Grouping had no main effect on achievement nor was there an interaction with instructor.
2. The retest option resulted in higher achievement.
3. Grouping seemed to be more effective for high ability students than for low ability students.
4. On basic objectives the whole class method was better than the grouping method.
5. Attitudes towards mathematics improved but there was no treatment effect.
6. There was no change on test anxiety.
7. None of the treatments had any differential effect on the drop-out rate.

Metfessel, Newton S. The Saugus Experiment in Multi-Grade Grouping. California Journal of Educational Research, Vol. 11 (1960).

One hundred twenty nine students in grades four, five and six were used in this study.

There were two classes in each of these grade levels. Three new classes were formed by mixing together one class from each of the three grade levels and randomly assigning these students to three new classes. The new classes contained approximately the same number of students from each of the three grades.

The other class at each of the three grade levels served as controls.

Standardized tests were administered at the end of the first year and again at the end of the second year for those students still in the school.

Students in the experimental and in the control classes were matched on sex, grade level, IQ, and age. There were six matched pairs of fourth graders, six matched pairs of fifth graders, and eleven matched pairs of sixth graders.

Findings:
1. The only significant difference at the end of the first year was that the fifth grade control students exceeded the fifth grade experimental students on arithmetic fundamentals.
2. This difference disappeared the next year when the fifth graders had become sixth graders.

Mikkelson, James E. An Experimental Study of Selective Grouping and Acceleration in Junior High School Mathematics. Doctoral dissertation, University of Minnesota (1962).

About seventy seventh and seventy eighth grade students were used in this study. The study was carried out twice, in two successive years.

In one junior high school, the top seventy (in mathematics) seventh graders (approximately the top 20%) were randomly split into two halves. One half made up an experimental, homogeneously grouped class. The other half was distributed, with the remaining students, into half the heterogeneously grouped classes. The same procedure was used the second year.

All seventh grade classes, including the experimental one, received the same curriculum each year. In the first year the curriculum was traditional and in the second year it was modern.

A standardized arithmetic reasoning test was administered as a pre- and a posttest each year. A comparison test was also used, as well as a local final test. An attitude inventory was used at the end of the year. An IQ test was also used at the beginning of the year.

The experimental eighth grade class was selected each year as was the seventh grade class. The control high ability students were distributed among half the regular classes.

The eighth grade experimental classes used acceleration. They studied the usual first year algebra course.

The control students were placed in an algebra class the next year with the teacher of the experimental class.

Standardized arithmetic tests were used at the beginning and end of the year.

Standardized algebra tests were used in the middle and at the end of the algebra course.

Findings:
1. Seventh grade gains were not significantly different for homogeneously and heterogeneously grouped high ability students in either year.
2. Attitudes did not differ across treatments.
3. Students not in the high ability group gained about the same whether or not there were high ability students in their classes.
4. Eighth grade results for the first year were the same for arithmetic as for the seventh grade on standardized tests. On a teacher-made test the control students scored higher the second year.
5. The seventh grade treatment had no effect on the eighth grade gains.
6. The eighth grade experimental gains (on eighth grade curriculum questions) were less than those for the controls.

7. The control students did better in algebra in grade nine than the experimental students in grade eight in both years, but both groups scored high in comparison to the norming groups for the tests.
8. The first year experimental grade eight students did as well on an algebra retest, after taking geometry in grade nine as the experimental students did after taking algebra in grade nine.

Millman, Jason; Johnson, Mauritz, Jr. Relation of Section Variance to Achievement Gains in English and Mathematics in Grades 7 and 8. American Educational Research Journal, Vol 1 (1964) pp. 47-51.

This study used 78 seventh grade and 90 eighth grade classes.

A standardized test was administered to seventh grade classes in New York State in the 1957-58 school year and to eighth grade classes the following school year. In 78 cases the tests were administered early in the school year and in 90 cases late in the school year.

For each class, the mean score on the arithmetic section of the first administration of the test was computed and also the standard deviation. The first was used to divide the classes into three ability levels and the second to divide them again into three levels of variability.

For each of the cells, the mean score on the second administration of the arithmetic test was computed.

Findings:
1. Except perhaps for the high ability students, there seemed to be no relationship between class variance and achievement.

Morgenstern, Anne. A Comparison of the Effects of Heterogeneously and Homogeneous Ability Grouping on the Academic Achievement and Personal-Social Adjustment of Selected Sixth-Grade Children. Doctoral dissertation, New York University (1963).

There were forty two sixth graders in one school and seventy seven in a second school.

In the smaller school the students had been homogeneously grouped, starting with grade four. In the other school heterogeneous grouping had been used.

A standardized test battery had been administered at the beginning of the fourth grade. It was repeated at the end of the sixth grade.

Some personality measures were also taken.

Findings:
1. There were no significant differences due to grouping on arithmetic computation or on arithmetic reasoning. (The experimental students gained more than the control students in language scores).
2. There was no interaction with IQ.
3. There were no significant differences in personal adjustment due to grouping.

Mortlock, Roland S. Provision for Individual Differences in Eleventh Grade Mathematics Using Flexible Grouping Based on Behavioral Ojectives: An Exploratory Study. Doctoral dissertation, University of Michigan (1969).

Two classes of eleventh grade students totaling thirty six altogether were used in this study. A similar set of students who had taken the same course the previous year was used as a control group.

The investigator wrote behavioral objectives at three levels: basic, intermediate, and advanced. These objectives were given to the students at the beginning of each new unit. The teacher then taught the unit to the class as a whole, aiming at the intermediate level objectives. Then a test was given and the class was grouped on the basis of the test. One group worked on unachieved basic objectives, another group worked on these and also on unachieved intermediate objectives. The top group also worked on advanced objectives.

During this second phase of the unit the teacher worked with the separate groups part of the time and the students worked individually the rest of the time.

A standardized algebra test was administered at the beginning and at the end of the year. A teacher constructed final examination which had also been used the previous year was given at the end of the experiment. An attitude inventory and a test anxiety scale were used as pre- and post-tests.

Findings:

1. On the final examination the only significant difference was that the control students did better than experimental students on advanced level objectives.
2. On the standardized algebra test there was no significant difference.
3. Student reaction to the teaching procedure was favorable
4. There were no changes in attitudes toward mathematics
5. Test anxiety decreased.

Peterson, Richard L. An Experimental Study of the Effects of Ability Grouping in Grades Seven and Eight. Doctoral dissertation, University of Minnesota (1966).

One hundred fifty two seventh graders and one hundred sixty five eighth graders were used in this study.

The students at each grade level were divided into three ability levels using IQ and a standardized achievement test battery. Each of the three levels was then randomly divided into two sets and each of these was subdivided into three subsets. At each grade level the three ability subsets in the first half of the population formed three homogeneously grouped classes. In the other half the three subsets were mixed together to form three heterogeneously grouped subsets.

Teachers were assigned randomly to classes and were asked to adjust their instruction to the ability of the class.

Standardized tests were administered at the beginning and at the end of the school year. An attitude inventory was given both to students and to the teachers.

Findings:
1. For the high ability students, there were no significant differences due to grouping on arithmetic computation, arithmetic concepts, arithmetic applications at either of the two grade levels.
2. For middle ability students, the seventh grade control heterogeneously grouped students learned significantly more than the experimental students on all three measures. In grade eight they did better than the experimental students on arithmetic computation and arithmetic concepts.
3. For the low ability students, there was no significant effect due to grouping on any of the three measures for the seventh graders, but for the eighth graders the controls did better than the experimental students on both arithmetic computation and arithmetic applications.
4. Middle level, grade seven experimental students showed no significant gain over the year on arithmetic applications.
5. Both middle and low level experimental students in grade eight showed no significant gain on arithemtic applications.
6. Students were relatively happy with whichever grouping system they were involved in.
7. Teachers generally favored homogeneous grouping.

Pinney, Grant C. Grouping by Arithmetic Ability--an Experiment in the Teaching of Arithmetic. The Arithmetic Teacher, Vol. 8 (1961).

Sixth grade students, about 60 in number, were studied. They were in the same school.

The students were divided into a fast and a slow class on the basis of previous arithmetic achievement. The instruction in each class was tailored to the ability level of the class. Standardized tests were given at the beginning and the end of the school year.

Findings:
1. The growth in arithmetic reasoning was from grade level 4.7 to grade level 6.8, in concepts from 5.1 to 7.3, and in computation from 5.0 to 7.3. (Separate statistics for the two classes are not provided.)
(The growth in spelling, to which no special attention was paid, was from 4.6 to 7.3.)

Provus, Malcolm M. Ability Grouping in Arithmetic. Elementary School Journal, Vol. 60 (1959-60) pp. 391-398.

Eleven intermediate grade classes formed the experimental group while eight classes served as a control group.

Experimental students were assigned to three ability levels for grades four and five and to two levels at grade six, and were grouped homogeneously.
Ability was defined as score on the arithmetic concepts scale of a standardized mathematics test, administered as a power test.
The criterion measure was the gain on this scale at the end of the year.
Another achievement test had been administered at the end of the previous year. It was readministered at the end of the school year.
An end of year test battery provided by the publisher of the text series used by the school system was also administered at the end of the school year.

An attitude inventory was administered at the beginning and the end of the year.

Findings:
1. High ability students scored significantly higher in homogeneously grouped classes than in heterogeneously grouped classes.
2. For middle and low ability students the two groupings were not statistically different.
3. Control students at the end of the year had about the same scores on the readministered test as corresponding students had the previous year. (If there was a Hawthorne effect, it did not affect the control students.)
4. On the readministered test, end of year scores were better than previous year scores for grades four and five, but not for grade six. (Ceiling effect?)
5. On the text series tests, at least one grade level of experimental students scored significantly higher than control students on each of the three scales.
6. No significant attitude changes were found.

Putbrese, Larry M. An Investigation into the Effect of Selected Patterns of Grouping upon Arithmetic Achievement. Doctoral dissertation, University of South Dakota (1971).

Three classes of fourth grade students were used in this study. All three classes were in the same school.

One class used the normal whole group instructional procedure. In the second class the students were grouped at three ability levels and the teachers spent one third of the class time with each group. Individualized instruction was used in the third class.

The total set of fourth grade students in the school was randomly assigned to these three classes.

Standardized tests were given at the beginning and the end of the year. In addition a modern mathematics test was administered at the beginning of the following school year.

Findings:
1. There was no significant difference between the achievement of the three classes over the school year.
2. There was no significant difference on retention over the summer.
3. The whole group class scored at a grade level significantly higher at the end of the summer on a modern mathematics test than it had on the original test at the end of the school year.

Savard, William G. An Evaluation of an Ability Grouping Program. California Journal of Educational Research, Vol. 11 (1960) pp. 56-60.

The students in this study were in grades 5 through 8. About 150 were included in the analysis, a sample drawn from a population of about 1200 students.

"Limited-range" grouping was used in one school district. The membership of a class consisted of average children plus a small group of above-average or below-average children, the respective classes being designated as "upper-range" or "lower-range".

In grades 4, 5, and 6, the primary criterion for grouping was a reading score. In grades 7 and 8, it was the average score on a standardized achievement battery.

For each student, gain during the year was compared with gain during the previous year, when heterogeneous grouping had been used.

Findings:
1. For upper-range classes, there was a significant increase in gains in arithmetic reasoning, but not in computation. For lower-range classes, there was a significant increase on both measures.
2. Gains were more pronounced for those with lower IQ.

Savard, William G. An Evaluation of the Second Year of an Ability Grouping Program. California Journal of Educational Research, Vol. 12, (1961) pp. 62-66.

The students in this study were those involved in limited-range classes the previous year.

The gains during the second year were compared with gains made two years previously.

Findings:
1 For arithmetic computation, gains during the year were significantly larger than during the earlier year for all but the students in upper-range classes for two years.
2. For those students in an upper-range class the first year but transferred to a lower-range class the second year, the arithmetic reasoning gain was significantly less. For all other students, there was no significant difference.
3. As in year 1, lower IQ students made greater gains.

Sawchuk, T. J.; Black, D. B. The Influence of Homogeneous Grouping on Teacher Marks in the High School. Alberta Journal of Educational Research, Vol. 7 (1961) pp. 156-170.

Two high schools were used in this study. The number of tenth grade students in one was 248 and in the other 197.

Students in the smaller school were randomly assigned to classes. In the larger school, ability grouping was used. Aggregate stanine scores on eight achievement tests given at the end of grade nine were used for this purpose. There were nine such classes.

Students in the smaller, heterogeneously grouped, school were divided into imaginary classes, one parallel to each class in the larger school, using the same procedure as was used in the larger school.

Teacher assigned grades were recorded for each student on seven subjects at the end of the tenth grade.

Findings:
1. In ninth grade mathematics, the nine pairs of classes were comparable. (The pairing had been done on the basis of aggregate ninth grade results.)
2. Overall achievement in mathematics was not significantly different between the two schools.
3. In four of the nine comparisons, the heterogeneously grouped classes exceeded the homogeneous classes in mathematics grades. For the other five cases the differences were not significant.

(Note: No achievement scores for the end of the tenth grade were collected, so it was not possible to find out how the grade distribution varied over the two grouping methods.)

Schmid, John A. A Study of the Uses of Sociometric Techniques for Forming Instructional Groups for Number Work in the Fifth Grade. Doctoral dissertation, University of Maryland (1960).

Four fifth grade classes were used in this study.

In the two control classes teachers formed intraclass groups when and as they pleased. In the experimental classes interclass groups for each unit were formed on the basis of a sociometric test.

IQ and reading scores for the four classes were similar. A standardized achievement test was administered at the beginning and end of the school year.

Findings:
1. The experimental classes achieved significantly more than the control classes on arithmetic reasoning and also on arithmetic computation.
2. Sociometric choices for mathematics and for recess were quite similar and were stable over time.

Schrank, William R. Relationships Between Ability Grouping and Academic Achievement in the Mathematics Course at the United States Air Force Academy Preparatory School. Doctoral dissertation, Texas A & M University (1967).

This study involved 420 cadets at the U. S. Air Force Academy.

At the Academy, ability grouping is normally used. In this experiment, a computer error assigned students essentially randomly to sections of a college freshmen-level mathematics course. The experiment lasted for three months.

The students were not told that ability grouping was not being used. However, the teachers were told what had happened.

Two examination grades and two progress report grades were obtained for each student. Means for each section were computed, and differences between the means of sections with different ability labels were compared.

Findings:
1. Of 257 comparisons, only 20 differences were significant. Most of these were due to poor performance of one section on one test and poor performance of another section on another test.

Schrank, Wilburn R. The Labeling Effect of Ability Grouping. The Journal of Educational Research, Vol. 62 (1968) pp. 51-52.

A brief reveiw of part of his dissertation.

Schrank, Wilburn R. A Comparison of Academic Achievement in Mathematics of Ability Grouped Versus Randomly Grouped Students. The Journal of Educational Research, Vol. 62 (1968) pp. 126-129.

A brief review of his dissertation.

Schrank, Wilburn R. A Further Study of the Labeling Effect of Ability Grouping. Journal of Educational Research, Vol. 63 (1970) pp. 358-360.

Two hundred and five cadets at the United States Air Force Academy Prepatory School were the subjects of this experiment.

Using scores on standardized mathematics tests, the cadets were divided into two equivalent sets. One set was grouped randomly for the first half of the year and by ability for the second half of the year. The opposite order was used for the other set.

Resectioning was carried out at the end of each unit of the course, randomly or by results on the unit test.

No instructor knew how his section had been grouped.

For half the students the curriculum was adjusted to the ability level.

The criterion variable was the assigned grade at the end of the semester.

Findings:

1. Random grouping was better than ability grouping when there was no curriculum differentiation.
2. Ability grouping was better than random grouping when curriculum differentiation was used.
3. When random assignment was used but when students or instructors thought that ability grouping was being used, then a label indicating high ability resulted in higher grades.

Schrank, Wilburn R. Academic Stimulation of Mathematics Pupils from their Classroom Association with Brighter Pupils. The Mathematics Teacher, Vol. 62 (1969) pp. 473-475.

A brief account of the preceding study.

Smith, William M. The Effect of Intra-Class Ability Grouping on Arithmetic Achievement in Grades Two through Five. Doctoral dissertation, Louisiana State University (1960).

The study involved one hundred ninety students in grades two through five.

Thirty two teachers were paired on teaching ability. One of each pair was then chosen at random for the experimental treatment.

Students within classes of paired teachers were then paired on the basis of previous arithmetic achievement and on IQ. There were ninety five such pairs scattered through the four grades.

In the experimental classes the students were grouped into three ability levels on the basis of a standardized test. At least 75% of the class time was spent working within the ability group.

The higher ability groups were allowed to proceed faster than the lower ability groups.

The control students were not grouped in any way.

A standardized test was administered at the end of the semester.

Findings:

1. The experimental students did better than the control students on computation in grades 2, 4, and 5.
2. The experimental students did better than the control students on problem solving in grade 2.

Sommers, Mildred E. A Comparative Study of Two Grouping Procedures in the Junior High School on Measures of Ability and Achievement in Mathematics and English. Doctoral dissertation, Michigan State University (1960).

Two hundred ninety one eighth grade students formed the experimental group for this study. They were all in the eighth grade. Two hundred thirty six students who had been in eighth grade two years previously formed the control group.

During the year that the control students were in grade eight, classes had been set up heterogeneously. For the experimental year students were grouped for English as well as mathematics on the basis of previous achievement. In each ability group students were encouraged to proceed at their own rate and the work was adjusted to the ability level. Standardized pre- and posttests were used.

Findings:

1. Although the control group was higher than the experimental group on numerical ability the experimental group scored significantly higher than the control group on arithmetic fundamentals at the end of the year.
2. The control group scored higher than the experimental group on arithmetic reasoning but the difference was not significant.

Steffani, Ronald R. Some Effects of Grouping by Subject Matter Major on Student Performance in College Calculus. Doctoral dissertation, Oregon State University (1970).

The experimental students in this study were the 41 out of 68 who completed a first year calculus course and had complete data on the covariates and the criterion tests. The control group consisted of the 47 students who completed the course the year before.

Three professors team-taught the experimental group, usually twice a week. The other two days of each week were spent in small groups. One group contained the biology, social science and business majors. A second contained the mathematics majors, and a third contained the physics, chemistry, and engineering students.

The discussions and assignments in the small groups were tailored to the interests of the students in that group.

A standardized calculus achievement test, an applied problem solving test, and an attitude inventory were administered at the end of the year. High school grade point average was used as a covariate.

Findings:
1. On the standardized calculus test the experimental students scored higher ($p < .10$) than the control students.
2. On the applied problem solving test the experimental students scored higher ($p < .20$) than the control students.
3. There was no significant difference in attitudes between experimental and the control students.

Steffani, R. R. Grouping by Academic Major in College Calculus. American Mathematical Monthly, Vol. 67 (1960) pp. 1135-1138.

A brief account of his dissertation.

Stern, Alfred M. Intraclass Grouping of Low Achievers in Mathematics in the Third and Fourth Grades. Doctoral dissertation, University of California, Los Angeles (1972).

This study involved twenty six third grade classes and twenty six fourth grade classes and concentrated on 138 low achieving third graders and 170 low achieving fourth graders in these classes.

The low achievers were identified on the basis of ratings given by the teachers who had taught the students the previous year. A mathematics test was also taken into account.

Roughly a third of the low achievers at each grade level were grouped within their classrooms and used a new (1970) elementary school text which had been prepared with the low achiever in mind. Another third were grouped within the classroom but did not use the special text. The final third served as the control group.

An attitude test was administered at the beginning and end of the school year as was also a standardized test. A special test prepared by the investigator to take into account the content of the special text was also administered at the beginning of the year and the end of the experiment, which lasted six months.

Teacher attitudes toward teaching mathematics and teacher experience were recorded.

Findings:
1. When pretests scores were taken into account there was no significant difference between the effects of the three treatments on student attitudes for third graders, but in the fourth grade the low achievers using the new text had better attitudes than the control students.
2. When pretest scores were taken into account there were no significant differences in arithmetic comprehension or arithmetic computation between the three treatments.
3. When pretests scores were taken into account there was no significant difference between the three treatments on the special test.
4. The teacher variables were not significantly related to student achievement.

Stevenson, Robert Louis. The Achievement Gains in Mathematics of Seventh Grade Pupils when Achievement Grouping and Flexible Scheduling are Employed in a Team Teaching Program. Doctoral dissertation, New York University (1966).

About one hundred forty seventh grade students were involved in this study.

The course was divided into ten units. At the beginning of each unit a diagnostic test was administered, after which the students were split into three levels; enriched, reinforced, and basic.

Team teaching and flexible scheduling were utilized.

A "mathematics ability" score was computed for each student on the basis of a standardized test, final grade in the sixth grade, and teacher recommendation.

A standardized test was administered at the beginning and end of the study.

Findings:

1. There was a positive correlation between mathematics ability and group placement.
2. Placement level changed for around forty percent of the students at the beginning of each unit. However, the placement levels were not independent of previous placements.
3. There was no significant difference on arithmetic computation or arithmetic applications between those who changed levels three or more times and those who changed two or fewer times. However, those who changed fewer were better than those who changed more often on arithmetic concepts.
4. There was a higher correlation between average placement level and achievement gains than between mathematical ability and achievement gains.
5. Over the entire student population there were significant gains over the school year on each of the three arithmetic scores.
6. About half the students had a significant gain on each of the three scores. The other half of the students had about the same ability but a lower average placement level.

Svensson, Nils-Eric. Ability Grouping and Scholastic Achievement. Report on a Five Year Follow-up Study in Stockholm. Stockholm Studies in Educational Psychology, No. 5, Almquist & Wiksell. Uppsala (1962).

The study involved all (except mentally retarded) students who were in the fourth grade in Stockholm in the 1954-55 academic year. There were a total of 341 classes with approximately 11,000 students at the beginning of the study.

On the north side of the city students were in heterogeneously grouped classes for the first four grades and then were placed in "plus-select" or "minus-select" classes on the basis of ability.

On the south side of the city students stayed in heterogeneously grouped classes till the end of sixth grade at which time they were separated into the two streams.

Socioeconomic data was gathered as well as IQ, reading, writing, and mathematics pre-scores in grade four. Tests were administered in the middle of grade 6 and at the end of each of grades 7, 8, and 9.

the plus-select students in each of the two systems were compared at three separate SES levels. The same was done for the minus-select students.

Findings:

1. The time of selection had no effect on mathematics scores either for the plus-select students or the minus-select students (this was not true for English).

Thelen, Herbert A. Classroom Grouping for Teachability. John Wiley & Sons, Inc., New York (1967).

Students involved in this study were in grades eight through eleven. Fifteen teachers were involved, some teaching geometry or advanced algebra.

A student personality inventory was used which assessed those student personality characteristics which were thought to be related to "teachability".

In the spring preceding the experimental year each teacher designated some of his current students as being "successful" and another set as being "unsuccessful". These students were administered the "personality assessment" battery and a scoring key was worked out for each teacher (different teachers had different views on "teachability"). These scoring keys were used during the following summer to select for each teacher an experimental class for the next year containing only students that the teacher thought would be easily teachable.

A control group was selected for each teacher by the school administrator using normal procedures.

An appropriate achievement test was administered at the beginning and again at the end of the year. Achievement was defined to be gain on the test. Teacher assigned grades were also recorded.

Findings:

1. Teachability scores were positively correlated with teacher assigned grades.
2. Teachability scores did not correlate with achievement.

Thelen, Herbert A. Grouping for Teachability. Theory into Practice, Vol. II (1963) pp. 31-39.

A brief discussion based on the above book.

Tuckman, Bruce W.; Bierman, Milton. Beyond Pygmalion: Galatea in the Schools. Paper presented at AERA, New York, February (1971). ED 047 077

About two hundred junior high school students were involved in this study.

The students had been assigned to ability levels, high, medium, and low, largely on the basis of teacher recommendations. The assignment was done independently for different school subjects.

Half of the students at each of the two lower ability levels were assigned to the experimental condition and were notified that they had been reassigned by the guidance department to the next higher level. (It was hypothesized that such an action would increase the experimental students self concept and would result in increased achievement.)

Standardized achievement and attitude tests were used.

Findings:

1. After one semester those who had been moved up did not do any better on a standardized test than those who had not.
2. Those moved from the medium to the high level got better teacher recommendations at the end of the semester but there was no difference on teacher assigned grades or on satisfaction.

3. Those moved from the low to the medium level got lower grades than those not moved and had less satisfaction.
4. In all cases, those moved up did not do as well as the students originally at the higher level.
(Note: The results for English and Social Studies were different.)

Wallen, Norman E.; Vowles, Robert O. The Effect of Intraclass Ability Grouping on Arithmetic Achievement in the Sixth Grade. Journal of Educational Psychology, Vol. 51 (1960) pp. 159-163.

Four sixth grade classes, with a total of 112 students, were used in this study.

There were two classes in each of two schools. Scores on a standardized test given at the end of grade five were used to separate the students into two matched and heterogeneous classes. Another form of the test was used at the beginning of the school year to check on the matching and to serve as a pre-measure.

In each school one teacher used a grouping procedure during the first half of the year and a non-grouping procedure for the rest of the year. The other teacher used these procedures in the opposite order.

In the grouping procedure, the students within the class were divided into four ability levels (using the most recent administration of the test). The teacher met with each group for at least three 20 minute periods a week. The rest of the time the students worked individually, but were encouraged to help each other. Text materials were appropriate to the ability level.

In the non-grouping procedure all students used the same text and were taught as a group. They had the same amount of time for individual work as did the experimental students.

A standardized test was administered at the middle and at the end of the school year.

Findings:
1. The treatment effect was not significant.
2. There was a significant teacher by treatment interaction.

Wardrop, James L., et al. Research and Development in R/I Units of Two Elementary Schools of Manitowoc, Wisconsin, 1966-67. ED 019 796

Eighty two third grade students were given a standardized arithmetic test.

To each student a score was assigned which was twice the arithmetic score added to the IQ score. These students were divided into thirds by means of this classifying score and then stratified by sex. One fourth of each sex in each ability level were assigned randomly to a control class. Thus there were four classes, low, medium, high, and control.

The low students were given manipulatives and went slower than the others. The high ability students were given some enrichment. There were four teachers. Each one taught each class for four weeks. At the end of the 16 weeks a standardized test and a home made test were administered.

Findings:
1. The average ability students in the homogeneous group performed better than those in the control group.
2. The low ability students in the homogeneous group performed less well than those in the control group.
3. For the high ability students neither form of grouping was better than the other.

West, Jeff; Sievers, Callie. Experiment in Cross Grouping. Journal of Educational Research, Vol. 54 (1960) pp. 70-72.

In this study the performance in mathematics of 28 secondary school students was examined.

These students were of high ability and had received instruction in language arts and in arithmetic, in grades three through six, in cross-grade homogeneously grouped classes. Other instruction was a heterogeneously grouped situation.

Teacher evaluations of student personality characteristics were obtained.

Scores on a standardized arithmetic test were also obtained.

Findings:
1. Leadership was high among the top 25% of these students but not among the top 10%.
2. In each of grades 7, 8, 9, and 10, these students scored above the national median in arithmetic.

Wilcox, John. A Search for the Multiple Effects of Grouping Upon the Growth and Behavior of Junior High School Pupils. Doctoral dissertation, Cornell University (1965).

Over one thousand eighth grade students in sixteen rural schools were in this study.

Grouping practices varied from school to school. A measure of the homogeneity of mental ages within a class was used to divide the schools into three groups: those using hetergeneous grouping, those using somewhat homogeneous grouping and those using quite homogeneous grouping. It was observed that there was little curriculum variation to take advantage of such grouping as there was.

Mathematics scores and mental ages were obtained from the school records. An attitude inventory and a critical thinking test were administered to all students.

Findings:
1. For the lower half of the IQ distribution, attitudes toward school were lower when classes were more homogeneous.
2. For upper IQ students mathematics achievement went up when classes were more homogeneous.
3. For upper SES students in the top half of the IQ distribution mathematics achievement went down when classes were more homogeneous.

4. For upper SES students in the upper half of the IQ distribution the more homogeneous the class the lower attitudes were toward school.
5. Grouping seemed to have no effect on critical thinking.

Wilcox, John. A Search for the Multiple Effects of Grouping Junior High School Pupils. Peabody Journal of Education, Vol. 41 (1964) pp. 216-225.

A summary of his dissertation.

Willcutt, Robert E. Ability Grouping by Content Subject Areas in Junior High School Mathematics. Doctoral dissertation, Indiana University (1967).

A total of 240 seventh grade students were used in this study.

Half of these students were assigned to four first-period classes and served as the controls. The remaining half was assigned to four second-period classes and were the experimental students.

The controls were assigned heterogeneously. The classes were independent, no student being moved from one class to another. Normal procedures were used.

The experimental students were assigned on the basis of a proficiency test given before each new subject matter unit, of which there were eight. There were two medium ability classes for each unit, one high ability and one low ability.

The teachers of the control classes also taught the experimental classes. They rotated the ability levels they taught as the year progressed.

For the experimental classes, the teachers were encouraged to differentiate the curriculum to adapt it to the ability level of the class.

Two standardized tests were administered at the beginning of the school year. Students in the control and experimental classes were matched on the basis of these tests, 78 matched pairs being designated.

Two other standardized tests and an original test were administered at the end of the school year. The last measured computation, structure, and the content of mathematics covered during the year.

An attitude inventory was administered at the beginning and the end of the year.

Findings:
1. Periodic regrouping, on the basis of prognostic tests, did not result in significant changes in achievement.
2. The attitudes of the experimental students improved more than those of the control students.

Willcutt, Robert E. Ability Grouping by Content Topics in Junior High School Mathematics. The Journal of Educational Research, Vol. 63 (1969) pp. 152-156.

A brief account of his dissertation.

Willcutt, Robert E. Individual Differences--Does Research Have any Answers for Junior High Mathematics Teachers? School Science and Mathematics, Vol. 69 (1969) pp. 217-225.

A discussion based on his dissertation.

Yerry, Marie J.; Henderson, Edward. Effects of Interage Grouping on Achievement and Behavior. End of Year Report. Palinedge Public Schools, Bethpage, New York (1964). ED 037 802

About five hundred elementary school students formed the experimental group for this study. About an equal number in another school served as the controls.

In the experimental school 22 classes were formed which cut across normal grade levels and thus were artifically heterogeneous. Thus, for example, there were six classes containing students that would normally be in grade one and other students that would normally be in grade two. There were three classes containing students from grade 1 through grade 3, etc.

Students in these classes worked in small groups or individually. The small groups were not always based on ability but could be formed for various reasons.

A standardized achievement test was administered at the beginning and again at the end of the year. An anxiety test was used both pre- and post and a sociometric friendship test was administered.

Findings:

1. On arithmetic there was no overall significant difference between the two schools. However, in grades 1 and 5 the experimental students scored slightly higher than the control students.
2. Second graders gained more if they were with first graders rather than with third graders. However, for fifth graders it did not matter whether they were with younger or older students.
3. The experimental treatment had no significant effect on the amount of friendship choices.
4. The experimental treatment had no significant effect on anxiety.

REVIEWS

Ekstrom, Ruth B. Experimental Studies of Homogeneous Grouping: A Critical Review. School Review, Vol. 69 (1961) pp. 216-226.

An Account of pre-1960 studies.

Findley, Warren G.; Bryan, Miriam M. Ability Grouping: 1970--Common Practices in the Use of Tests for Grouping Students in Public Schools, 1970. ED 048 381

Findley, Warren G.; Bryan, Miriam M. Ability Grouping: 1970--II The Impact of Ability Grouping on School Achievement, Affective Development, Ethnic Separation, and Socio-economic Separation, 1970. ED 048 382

Findley, Warren G.; Bryan, Miriam M. Ability Grouping: 1970--III The Problems and Utilities Involved in the Use of Tests for Grouping Children with Limited Backgrounds, and Alternative Strategies to Such Grouping, 1970. ED 048 383

Findley, Warren G.; Bryan, Miriam M. Ability Grouping: 1970--IV Conclusions and Recommendations, 1970. ED 048 384

An extensive review of grouping, not restricted to mathematics.

Heathers, G. Grouping. Encyclopedia of Educational Research, Fourth Edition. pp. [illegible].

An extensive discussion.

Passow, A. Harry. The Maze of the Research on Ability Grouping. The Educational Forum, Vol. 26 (1961-62) pp. 281-288.

Yates, Alfred. Grouping in Education. John Wiley & Sons, New York (1966). 371.24 Y31

This report is of interest because it was based on an international conferenee.